



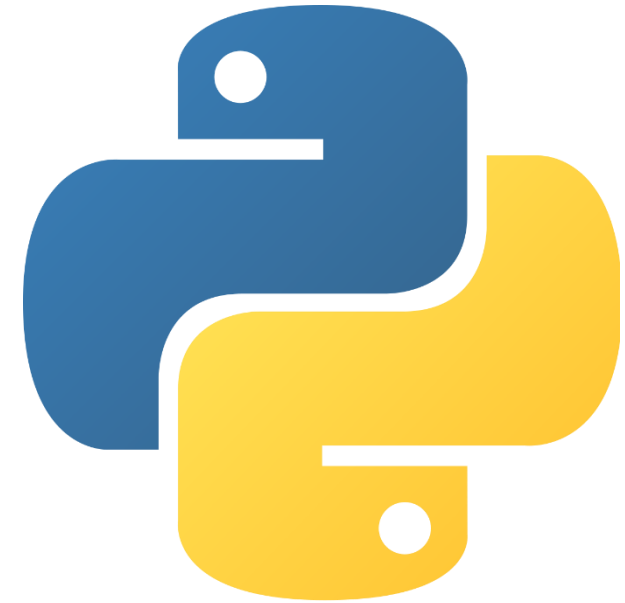
# DATA MINING TOOLS

# Outline

- Python
- Jupyter Notebook
- Colab

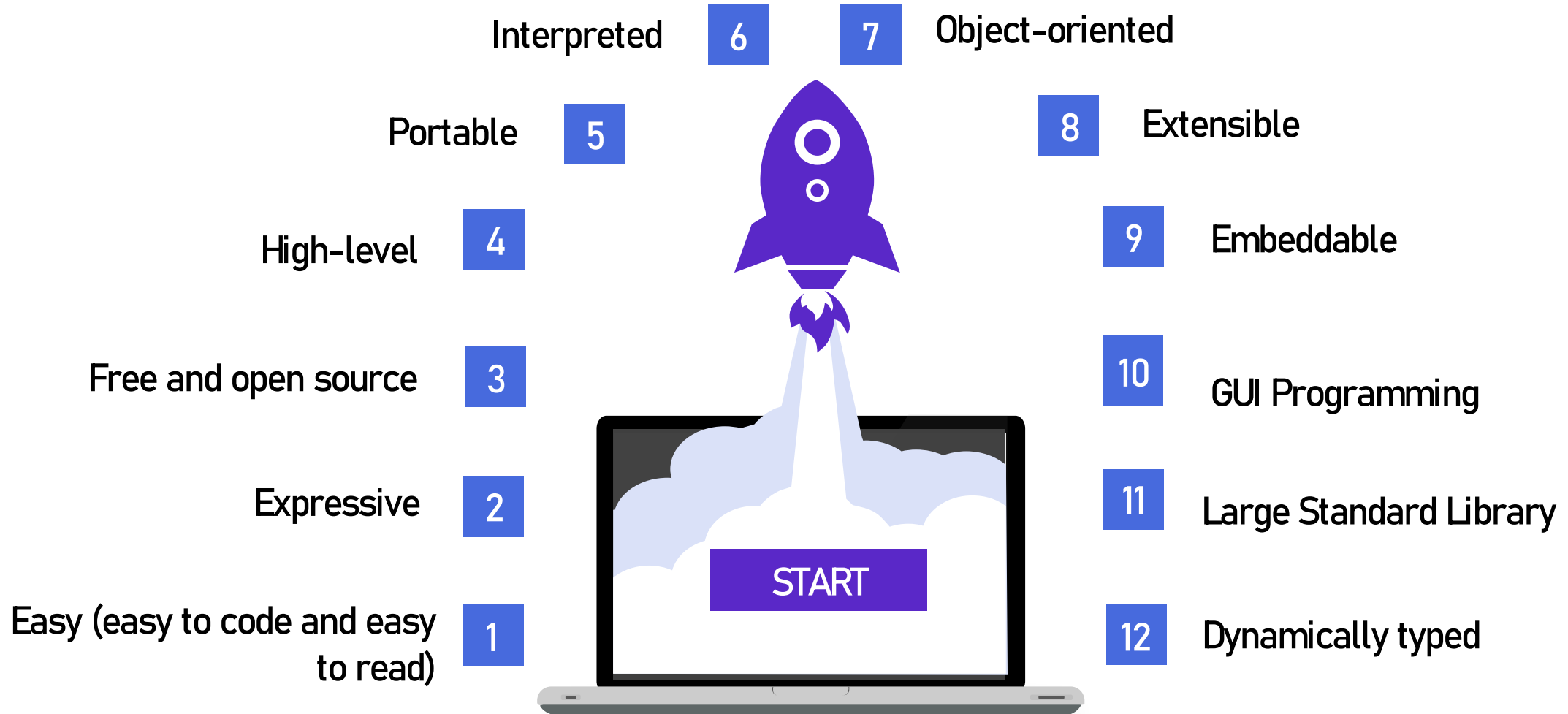
# What is Python

Python is an interpreted, interactive, object-oriented programming language. It incorporates modules, exceptions, dynamic typing, very high level dynamic data types, and classes. It supports multiple programming paradigms beyond object-oriented programming, such as procedural and functional programming. Python is portable: it runs on many Unix variants including Linux and macOS, and on Windows.



Source: <https://docs.python.org/3/faq/general.html#what-is-python>

# Reason to Use Python



# Kegunaan Python

- Implement Machine Learning (Give a computer the ability to learn)
- Perform Data Analysis using Python
- Build a website using Python
- Develop a game in Python
- Perform Computer Vision (Facilities like face-detection and color-detection)
- Enable Robotics with Python
- Perform Web Scraping (Harvest data from websites)
- Automate a web browser
- Perform Scripting in Python
- Perform Scientific Computing using Python
- Build Artificial Intelligence
- etc

# Top 10 Most Popular Programming Languages For 2021 and Beyond

Language Rank	Types	Spectrum Ranking
1. Python		100.0
2. C		99.7
3. Java		99.5
4. C++		97.1
5. C#		87.7
6. R		87.7
7. JavaScript		85.6
8. PHP		81.2
9. Go		75.1
10. Swift		73.7

Source:  
<https://becominghuman.ai/>

# Python & Installation



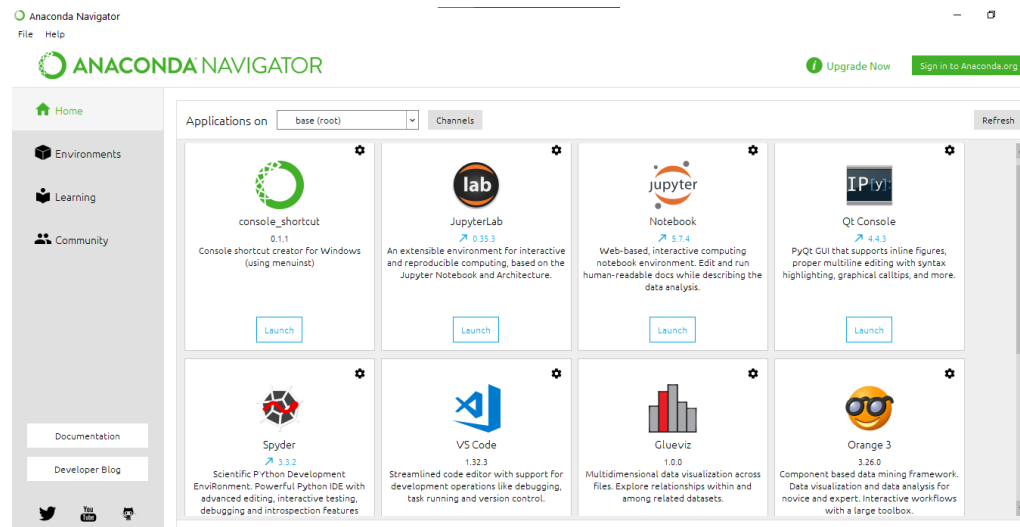
Can download from project site and install libraries individually

<https://www.python.org/downloads/>



Download python with Anaconda distribution

<https://www.anaconda.com/products/individual>



Anaconda is a free and open-source distribution of the programming languages Python and R. The distribution comes with the Python interpreter and various packages related to machine learning and data science.

# Jupyter Notebook

The **Jupyter Notebook** is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.



```
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3
In [9]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split

In [3]: import os
print(os.listdir("G:/My Drive/Pendidikan/Semester 2/ML FKG"))

['heart.csv']

In [6]: # We are reading our data
df = pd.read_csv("G:/My Drive/Pendidikan/Semester 2/ML FKG/heart.csv")

In [7]: # First 5 rows of our data
df.head()
```

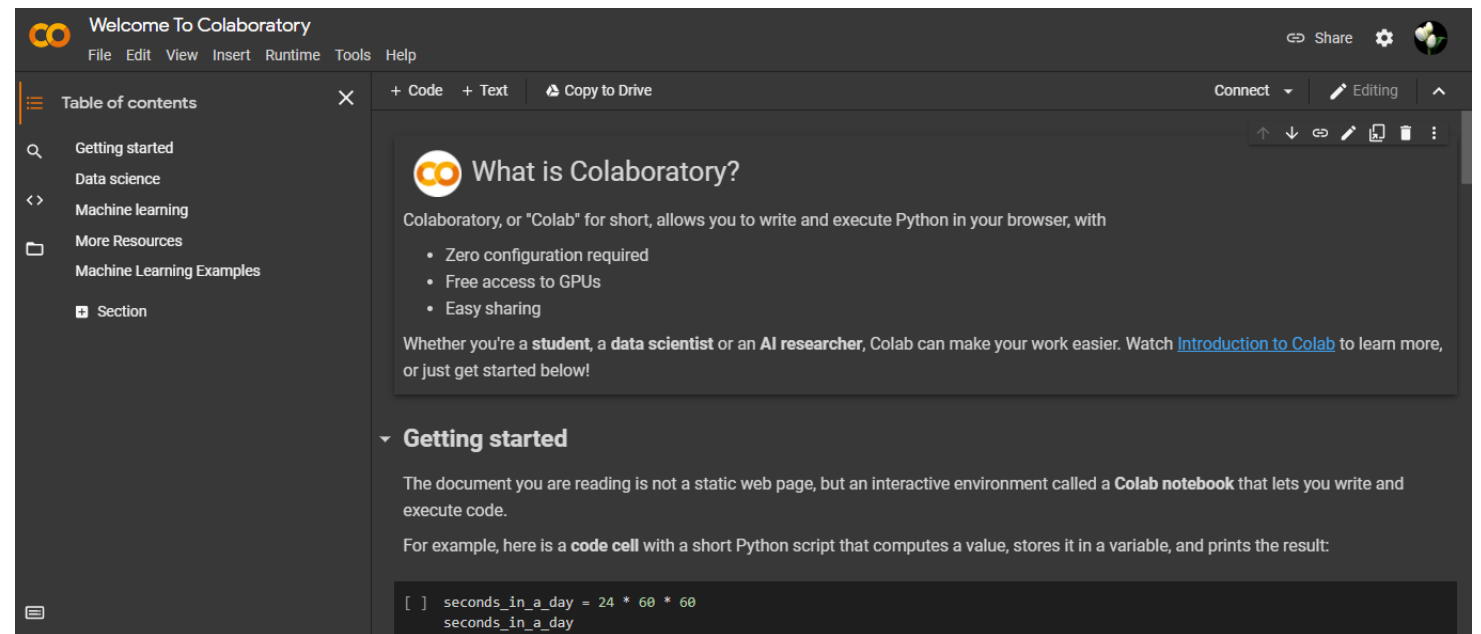
Source: <https://jupyter.org/>

# Colab

Colaboratory, or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs.



<https://colab.research.google.com/>



Source: <https://research.google.com/colaboratory/faq.html>

# Python vs other languages

## Python

```
print("Hello World")
```

## C

```
#include <stdio.h>

int main(void)
{
    printf("hello, world\n");
}
```

## C++

```
#include <iostream>

int main()
{
    std::cout << "Hello, world!\n";
    return 0;
}
```

## Java

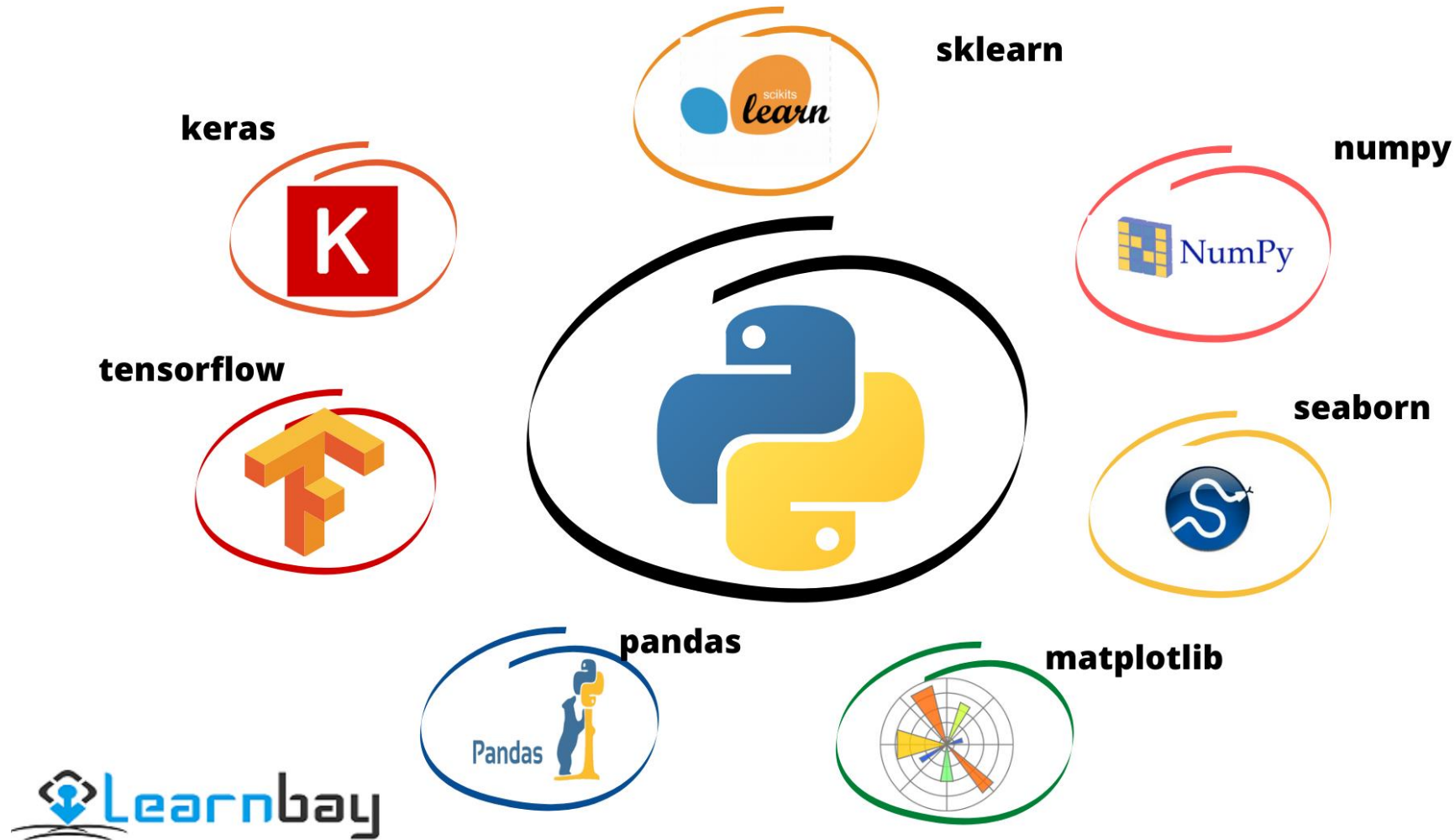
```
class HelloWorldApp {
    public static void main(String[] args) {
        System.out.println("Hello World!"); // Prints the string to the
        console.
    }
}
```

## Pascal

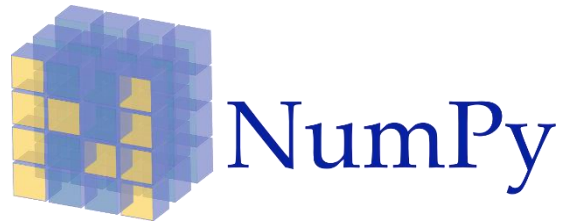
```
program HelloWorld(output);
begin
    Write('Hello, world!')
end.
```

Source: <https://excelwithbusiness.com/blog/say-hello-world-in-28-different-programming-languages/>

# Python Libraries



# Python Libraries



NumPy is a perfect tool for scientific computing and to perform from basic to advanced array operations.



Matplotlib is a standard library for data science that helps in generating data representation in 2D and 3D planes.

Pandas



Pandas has all the data manipulation, optimization, visualization, and data wrangling features.



Scikit learn (sklearn) is mostly used for data science projects (machine learning algorithms and data mining problems)

# Python Data Type

Name	Type	Description
Integers	int	Whole numbers, such as: <b>3 300 200</b>
Floating point	float	Numbers with a decimal point: <b>2.3 4.6 100.0</b>
Strings	str	Ordered sequence of characters: <b>"hello" 'Sammy' "2000" "楽しい"</b>
Lists	list	Ordered sequence of objects: <b>[10,"hello",200.3]</b>
Dictionaries	dict	Unordered Key:Value pairs: <b>{"mykey" : "value" , "name" : "Frankie"}</b>
Tuples	tup	Ordered immutable sequence of objects: <b>(10,"hello",200.3)</b>
Sets	set	Unordered collection of unique objects: <b>{"a","b"}</b>
Booleans	bool	Logical value indicating <b>True</b> or <b>False</b>

# Numeric

In Python, numeric data type represent the data which has numeric value.

**Integers** – This value is represented by *int* class. It contains positive or negative whole numbers (without fraction or decimal). In Python there is no limit to how long an integer value can be.

**Float** – This value is represented by *float* class. It is a real number with floating point representation. It is specified by a decimal point. Optionally, the character e or E followed by a positive or negative integer may be appended to specify scientific notation.

**Complex Numbers** – Complex number is represented by *complex* class. It is specified as *(real part) + (imaginary part)j*. For example –  $2+3j$

## Contoh:

### Numeric

```
▶ a = 5
print("Type of a: ", type(a))

b = 5.0
print("\nType of b: ", type(b))

c = 2 + 4j
print("\nType of c: ", type(c))
```

```
↳ Type of a: <class 'int'>
Type of b: <class 'float'>
Type of c: <class 'complex'>
```

# Boolean

Data type with one of the two built-in values, True or False. Boolean objects that are equal to True are truthy (true), and those equal to False are falsy (false). But non-Boolean objects can be evaluated in Boolean context as well and determined to be true or false. It is denoted by the class bool.

**Note** – True and False with capital 'T' and 'F' are valid booleans otherwise python will throw an error.

Contoh:

```

Boolean

[20] print(type(True))
      print(type(False))

      print(type(true))

<class 'bool'>
<class 'bool'>
-----
NameError                                Traceback (most recent call last)
<ipython-input-20-41b100f19035> in <module>()
      2 print(type(False))
      3
----> 4 print(type(true))

NameError: name 'true' is not defined
    
```

# String

In Python, Strings are arrays of bytes representing Unicode characters. A string is a collection of one or more characters. It is represented by *str* class.

Contoh:

```
String
▶ String = 'Welcome to the Python World'
print("Sentence: ")
print(String)
print(type(String))
↳ Sentence:
Welcome to the Python World
<class 'str'>
```

# Lists

- Lists are used to store multiple items in a single variable.
- List items are ordered, changeable, and allow duplicate values.
- List items are indexed, the first item has index `[0]`, the second item has index `[1]` etc.

Contoh:

```
[1] list1 = ["apple", "banana", "cherry"]  
print(list1)
```

```
['apple', 'banana', 'cherry']
```

```
▶ list2 = [1, 5, 7, 9, 3]  
print(list2)
```

```
[1, 5, 7, 9, 3]
```

# Tuples

- Tuples are used to store multiple items in a single variable.
- Tuple items are ordered, unchangeable, and allow duplicate values.
- Tuple items are indexed, the first item has index `[0]`, the second item has index `[1]` etc.

Contoh:

```
Tuples

[12] thistuple = ("apple", "banana", "cherry")
      print(thistuple)

('apple', 'banana', 'cherry')

▶ thistuple2 = (1, 5, 7, 9, 3)
   print(thistuple2)

(1, 5, 7, 9, 3)
```

# Sets

- Sets are used to store multiple items in a single variable.
- A set is a collection which is both *unordered* and *unindexed*.
- Sets are written with curly brackets.
- Set items are unordered, unchangeable, and do not allow duplicate values.

Contoh:

```
Sets

[13] thisset = {"apple", "banana", "cherry", "apple"}
      print(thisset)

      {'apple', 'cherry', 'banana'}

▶ thisset2 = {1, 5, 7, 9, 3}
   print(thisset2)

   {1, 3, 5, 7, 9}
```

# Dictionaries

- Dictionaries are used to store data values in key: value pairs.
- A dictionary is a collection which is ordered, changeable and does not allow duplicates.
- Dictionaries are written with curly brackets, and have keys and values.

Contoh:

```
Dictionaryes

▶ thisdict = {
  "brand": "Ford",
  "model": "Mustang",
  "year": 1964
}
print(thisdict)

{'brand': 'Ford', 'model': 'Mustang', 'year': 1964}
```



# Terima Kasih